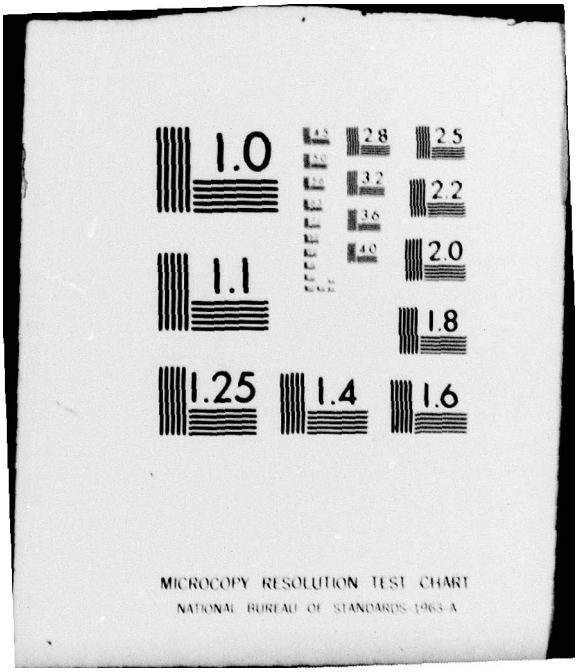
AD-A074 257 PURDUE UNIV LAFAYETTE IND DEPT OF STATISTICS F/G 12/1
PROBABILITY AND STATISTICS AND APPLICATIONS.(U)
JUL 79 L J GLESER, D S MOORE AFOSR-77-3291
UNCLASSIFIED AFOSR-TR-79-0961 NL

1 OF 1
AD A074257

END
DATE FILMED
10-79
DDC


1.0
1.1
1.25
1.4
1.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.8
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A074257

DDC FILE_COPY

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR-79-0961	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): PROBABILITY AND STATISTICS AND APPLICATIONS	5. TYPE OF REPORT & PERIOD COVERED: FINAL
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s): Leon Jay Gleser, David S. Moore	8. CONTRACT OR GRANT NUMBER(s): AFOSR-77-3291
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics and Purdue Research Foundation, Purdue University, West Lafayette, IN 47907	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, D.C. 20332	12. REPORT DATE: July 1979
	13. NUMBER OF PAGES: 7
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 11p.	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public release; distribution unlimited.

Final scientific rept. 1 Jun 77-31 May 79

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DDC RECEIVED SEP 25 1979 A

18. SUPPLEMENTARY NOTES

the following

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Minimax estimation, admissible estimation, vector parameter, multivariate normal discrete observations, multivariate linear model, ridge regression, Bayesian multiple comparisons, two factor designs, errors in variables, linear functional equations, large sample confidence regions, sequential Bayes multiparemeter regional estimation procedures, asymptotically pointwise optimal, (cont)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This final report reviews work by Drs. Leon Jay Gleser and David S. Moore, and various students, on the following topics: minimax estimation of the mean vectors of multivariate normal distributions and of the slopes of multivariate linear regression models when the error covariance matrix is unknown; minimax properties of generalized ridge regression estimators of regression coefficients in the classical linear model; improvements on inadmissible simultaneous estimators for the parameters of k discrete populations; Bayes multiple comparison procedures in two factor ANOVA designs; calculations of and properties of point and (cont)

DD FORM 1 JAN 73 1473 291730 UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19(Cont.) chi-square goodness of fit tests, censored data, chi-squared tests of multivariate normality.

20. (Cont.) interval estimators of slope parameters in multivariate errors in variables regression models (linear functional equations); asymptotic pointwise optimal and asymptotically optimal sequential Bayes regional estimators for functions of a vector parameter; chi-square goodness-of-fit tests for censored data and for multivariate normality; and the general theory and application of chi-square tests. A bibliography of publications, reports, and dissertations supported by the grant is provided.

FINAL SCIENTIFIC REPORT

on

PROBABILITY AND STATISTICS AND APPLICATIONS

Grant AFOSR-77-3291

Period: June 1, 1977 - May 31, 1979

Prepared by

Leon Jay Gleser, Principal Investigator

David S. Moore

Department of Statistics

and

Purdue Research Foundation

Purdue University

July 1979

FINAL SCIENTIFIC REPORT

on

"PROBABILITY AND STATISTICS AND APPLICATIONS"

Grant AFOSR-77-3291

June 1, 1977 - May 31, 1979

I. PERSONNEL SUPPORTED

Dr. Leon Jay Gleser, Principal Investigator
Dr. David S. Moore
Dr. George Casella (June 1 - August 1, 1977)
Dr. Daniel P. Milhalko (June 1 - August 1, 1977)
Mrs. Alice Sun Huang (September 1, 1977 - May 31, 1978)
Mr. Richard Sundheim (September 1, 1977 - August 1, 1978)
Mr. V. Ramaswami (September 1, 1977 - December 31, 1977)
Mr. Jiunn Fzon Hwang (September 1, 1978 - May 31, 1978)

II. SUMMARY OF RESEARCH AND OTHER ACTIVITIES

Numbered references (e.g.[3]) refer to the bibliography of publications, etc., in Part III of this report.

A. RESEARCH OF DR. LEON JAY GLESER

1. Improved Point Estimators of Vector Parameters

Let X be a random p-variate normally distributed vector having unknown mean vector μ and unknown covariance matrix Σ, and let W be independently distributed with a Wishart distribution with n degrees of freedom and expected value $n\Sigma$. In [3], which will appear in the July, 1979 Annals of Statistics, Dr. Gleser exhibits a broad class of estimators for μ each of which, under general quadratic loss, dominate the usual estimator X in risk ($p \geq 3$). A generalization of the problem of this paper to the case where k independent p-dimensional normally distributed random vectors $X_1, X_2, \ldots, X_k$ with unknown mean vectors $\mu_1, \mu_2, \ldots, \mu_k$ and common unknown covariance Σ are observed along with W is treated in [5], currently in preparation.

The technical report [1] by Dr. George Casella concerns the minimaxity of generalized ridge regression estimators for the parameters of the classical

univariate linear regression model. A revision of this report has been accepted for publication by the Annals of Statistics.

The dissertation [6] by Mr. Jiunn Fzon Hwang, written under the direction of Dr. James O. Berger and partially supported by the grant, concerns improved estimators of the components of a vector parameter $\underline{\theta} = (\theta_1,\ldots,\theta_p)$ of the joint distribution of p independent discrete observations $X_1,\ldots,X_p$, where the marginal distribution of X_i had the discrete exponential form $A_i(\theta_i)B_i(X_i)\theta_i^{X_i}$. The loss function for estimating $\underline{\theta}$ by $\underline{a} = (a_1,a_2,\ldots,a_p)$ is of the form $\sum_{i=1}^{p} c_i\theta_i^m(\theta_i-a_i)^2$. In Hwang's work, the problem of improving upon inadmissible estimators is reduced to the study of generalized difference inequalities. Typical difference inequalities, including those appearing for the Poisson and negative binomial distributions are presented and solved. Theorems are obtained which establish the inadmissibility of certain broad classes of estimators. Hwang's work generalizes previous research in this area by: (1) Use of a more generalized loss function, (2) illustrating how to improve upon biased (as well as unbiased) estimators, and (3) exhibiting broader classes of estimators which improve both upon the usual estimators and upon some "improved" estimators previously suggested in the literature. Mr. Hwang will receive his Ph.D. in August 1979 and will then become a faculty member at Cornell University.

2. Bayesian Multiple Comparisons in Factorial Designs

Mrs. Alice Sun Huang worked under Dr. Gleser's direction on the problem of finding Bayes rules for multiple comparisons of the cell means μ_{ij} in a balanced two factor analysis of variance design, generalizing the work of Duncan and Waller [Journ. Amer. Statist. Assoc. (1969), 1484-1503] for one-factor designs. Mrs. Huang found the form of the Bayes rules, and was working on algorithms to compute cut-off points for the decision rules when she elected to take a leave of absence to obtain industrial work experience.

3. "Errors-in-Variables" Multivariate Regression Models

Dr. Gleser extensively revised his paper [2] in response to comments of referees for the Annals of Statistics. New material was added to the paper relating to an extension of the "errors-in-variables" regression model to include an intercept, and to the question of whether the assumptions made in the paper about the form of the error covariance matrix could be relaxed.

The technical report [4] concerns algorithms for calculation of the maximum likelihood estimators of the model, and the lack of feasibility for simulations required to verify small-sample coverage properties of the large-sample confidence region proposed in [2] for the slope parameters of the model when the total number of variables exceeds 5. This paper was presented as an invited talk at the 1978 Eastern Regional Meeting of the Insitute of Mathematical Statistics at New Brunswick, New Jersey. No publication of this paper is planned. A first draft of a new technical report comparing three confidence intervals proposed in the literature for the slope parameter in the one-dimensional case (one independent and one dependent variable) is nearly complete. The conclusion of this new technical report is that none of the three interval estimation procedures is totally satisfactory in terms of coverage (large and small sample) and length.

4. Sequential Bayes Regional Estimation Procedures

The dissertation [15] by Mr. Richard Sundheim, written under the direction of Dr. Gleser, extends prior work of Gleser and Kunte [Ann. Statist. (1976), 685-711] concerning asymptotically pointwise optimal (A.P.O.) and asymptotically optimal (A.O.) sequential Bayes interval estimators in single-parameter models to the case of A.P.O. and A.O. sequential Bayes regional estimators for functions $g_i(\theta)$, $1 \leq i \leq r$, of a p-dimensional parameter vector $\theta (r \leq p)$. The dissertation also gives sufficient conditions under which it is possible to approximate the Bayes rules with rules of simpler functional form, while still maintaining A.P.O. and A.O. properties. These last results provide some answers to the question of how robust the Bayes procedures are to the choice of the prior distribution for θ. Mr. Sundheim will receive his Ph.D. degree in August 1979; he is currently a faculty member at Kansas State University.

5. Other Activities and Honors

Dr. Gleser has given invited talks on work supported by this grant at the 1978 Eastern Regional Meeting of the Institute of Mathematical Statistics and at Eastern Illinois University. He is currently a member of the Management Committee of The Journal of Educational Statistics, and has just ended a seven year tenure as Associate Editor of the journal Psychometrika. Dr. Gleser organized a session on multivariate analysis at the 1978 Central Regional Meeting of the Institute of Mathematical Statistics held at Lexington, Kentucky.

Mr. Richard Sundheim and Mr. Jiunn Hwang received the Irving W. Burr Award of the Department of Statistics at Purdue University in the years 1978 and 1979, respectively. This award is granted annually to the Ph.D. student in the department who has made, and promises to make, the greatest contribution to the department and the statistical profession.

B. RESEARCH OF DR. DAVID S. MOORE

1. General Chi-Square Statistics for Lack of Fit

Research continued on applications and extensions of the theory of general chi-square statistics for tests of fit, building on the earlier work of Moore and Spruill [Ann. Statist. (1975), 599-616].

In [8], Dr. Moore and Dr. Daniel Mihalko extended the Moore-Spruill Theory to the case of randomly censored data. Data of this type are common in life testing situations. Based on the general theory, explicit usable statistics for testing the fit of censored samples to the exponential, normal, uniform and Weibull families of distributions were derived. These statistics use convenient estimators of unknown parameters(the maximum likelihood estimator in cases where it has a closed form expression) and have chi-square limiting distributions under the null hypothesis. They are therefore easily computable and require only standard tables of critical points. At the suggestion of the referees, substantial additional theoretical work was done to extend the theory to include asymptotic behavior under contiguous alternatives. Thus extended, the paper has been accepted by the Annals of Statistics to appear in May, 1980.

An application of the Moore-Spruill Theory to testing multivariate normality was made by Moore and Stubblebine in [11]. Data-dependent cells are employed which are bounded by hyperellipses centered at the sample means and having shape determined by the sample covariance matrix. Such cells can be chosen to be equiprobable under the (composite) hypothesis of multivariate normality, so that the cell frequencies are immediately indicative of departures such as peakedness or heavy tails. Chi-square tests based on these cells have many convenient properties, which are studied and reported. These tests are then applied to test the multivariate normality of returns on common stocks, an assumption commonly made in the theory of financial markets. A large data base (midweek closing prices for 560 stocks over 501 weeks) is employed to study normality of returns and to obtain results and make suggestions on

practical use of the proposed tests.

In addition, final revisions were made on Dr. Moore's invited contribution [9] to the Studies In Statistics volume of the Mathematical Association of America. This volume has now appeared.

2. Other Activities and Honors

Dr. Moore is currently serving as chairman of the Satistical Education Section of the American Statistical Association, and as a member of the Special Papers Committee of The Institute of Mathematical Statistics. He was invited to contribute an essay on "Statistical Analysis of Experimental Data" to the volume Mathematics Today: Twelve Informal Essays (Springer-Verlag, 1979), a volume prepared for the Joint Projects Committee of the mathematics professional societies.

C. OTHER RESEARCH

Mr. V. Ramaswami did his dissertation research under partial support of both this grant and a previous grant (AFOSR-72-2350C) at Purdue University. Although he received his Ph.D. from Purdue University in August 1978, he did his work at the University of Delaware, under the direction of Professor Marcel Neuts and also under partial support of grant AFOSR-77-3236 at that university. Technical reports [12], [13], [14][1], the last two of which were written jointly with Professor Neuts, are based on Ramaswami's dissertation. Since these reports have been described in the annual report for Grant AFOSR-77-3236 by Professor Neuts, and copies have been submitted to the Air Force, it would be redundant to describe them here. However, they are listed in the bibliography of papers in Part III.

III. BIBLIOGRAPHY OF PUBLICATIONS, TECHNICAL REPORTS, AND DISSERTATIONS SUPPORTED BY THE GRANT

[1] Casella, George. Minimax ridge regression. Purdue Department of Statistics Mimeograph Series, No. 497. A revised paper has been accepted by the Annals of Statistics.

[1] In these reports Ramaswami acknowledges support from Grant AFOSR-72-2350C but mistakenly fails to acknowledge support of the present grant.

[2] Gleser, Leon Jay. Estimation of a linear transformation: Large sample results. Purdue Department of Statistics Mimeograph Series, No. 453. A revised paper has been accepted by the Annals of Statistics.

[3] Gleser, Leon Jay. Minimax estimation of a normal mean vector when the covariance matrix is unknown. Purdue Department of Statistics Mimeograph Series, No. 510. To appear in July 1979 Annals of Statistics.

[4] Gleser, Leon Jay. Calculation and simulation in errors-in-variables regression problems. Purdue Department of Statistics Mimeograph Series, No. 78-5. Invited paper given at 163rd meeting of the Institute of Mathematical Statistics, New Brunswick, New Jersey, June 1, 1978.

[5] Gleser, Leon Jay. Improved point estimators of the slopes in multivariate linear regression with unknown covariance matrix. Purdue Department of Statistics Mimeograph Series, No. 79-12.

[6] Hwang, Jiunn Fzon. Improving upon inadmissible estimators in discrete exponential families. Ph.D. dissertation, Department of Statistics, Purdue University, awarded August 1979.

[7] Mihalko, Daniel P. Chi-square tests of fit for Type II censored data. Ph.D. dissertation, Department of Statistics, Purdue University, awarded August, 1977. Revised as [8].

[8] Mihalko, Daniel and Moore, David S. Chi-square tests of fit for Type II censored data. Purdue Department of Statistics Mimeograph Series, No. 505. To appear in May 1980 Annals of Statistics.

[9] Moore, David S. Chi-square tests. In Robert Hogg (ed.) Studies in Statistics, 66-106. The Mathematical Association of America, 1978.

[10] Moore, David S. Chi-square techniques. In Ralph D'Agostino and M.A. Stephens (Eds.) Goodness of Fit Techniques. Marcel Dekker, to appear.

[11] Moore, David S. and Stubblebine, John B. Chi-square tests for multivariate normality, with application to common stock prices. J. Amer. Statist. Assoc., submitted.

[12] Ramaswami, V. The N/G/1 Queue and its detailed analysis. University of Delaware Department of Statistics and Computer Science Mimeograph Series, No. 78/1. April, 1978.

[13] Ramaswami, V. and Neuts, Marcel F. Some explicit formulas and computational methods for infinite server queues with phase type arrivals. University of Delaware Department of Statistics and Computer Science Mimeograph Series, No. 78/5. April, 1978.

[14] Ramaswami, V. and Neuts, Marcel F. A duality theorem for phase type queues. University of Delaware Department of Statistics and Computer Science Mimeograph Series, No. 78/11. June, 1978.

[15] Sundheim, Richard. Asymptotically optimal multiparameter sequential Bayes regional estimation procedures. Ph.D. dissertation, Department of Statistics, Purdue University, awarded August, 1979.